UNIVERSITY of York

This is a repository copy of *Utilitarianism without Moral Aggregation*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/175185/

Version: Published Version

Article:
Gustafsson, Johan orcid.org/0000-0002-9618-577X (2021) Utilitarianism without Moral Aggregation. Canadian journal of philosophy. pp. 256-269. ISSN 0045-5091

https://doi.org/10.1017/can.2021.20

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Canadian Journal of Philosophy (2021), 51: 4, 256–269
doi:10.1017/can.2021.20

ARTICLE

Utilitarianism without Moral Aggregation

Johan E. Gustafsson

University of Gothenburg, University of York, Institute for Futures Studies
Email: johan.eric.gustafsson@gmail.com

Abstract
Is an outcome where many people are saved and one person dies better than an outcome where the one is saved and the many die? According to the standard utilitarian justification, the former is better because it has a greater sum total of well-being. This justification involves a controversial form of moral aggregation, because it is based on a comparison between aggregates of different people's well-being. Still, an alternative justification—the Argument for Best Outcomes—does not involve moral aggregation. I extend the Argument for Best Outcomes to show that any utilitarian evaluation can be justified without moral aggregation.

Keywords: Moral aggregation; the Argument for Best Outcomes; combining claims; the number problem; utilitarianism

Is an outcome where many people are saved and one person dies better than an outcome where the one is saved and the many die? Most of us judge that the former is better. But what justifies this evaluation? The standard utilitarian answer is that it would be better if the many were saved, because the combined gain in well-being for the many if they were saved would be greater than the gain in well-being for the one if he or she were saved.[1] This form of utilitarianism justifies evaluations by

> *The Total Principle*: Outcome *X* is at least as good as outcome *Y* if and only if the sum total of well-being is at least as great in *X* as in *Y*.

The justification by the Total Principle is an example of moral aggregation, which some people, such as John M. Taurek and T. M. Scanlon, find problematic. Taurek, for example, complains that

> It is not my way to thinking of [the people who need help] as each having a certain *objective* value, determined however it is we determine the objective value of things, and then to make some estimate of the combined value of the [many] as against the one. (1977, 307)[2]

Scanlon is somewhat less clear, demanding that

> the justifiability of a moral principle depends only on various *individuals*' reasons for objecting to that principle and alternatives to it. (1998, 229)

An aggregate or sum of several individuals' reasons, however, still depends on those individual reasons. Yet, since Scanlon takes his demand to rule out justifications that appeal to a '*sum* of a certain sort of value' (230), he seems to have in mind a requirement that is, more or less, equivalent to Taurek's requirement.

[1]Timmons (2013, 117) and Portmore (2020, 5–6) put forward utilitarian accounts of rightness with this kind of total justification.

[2]See also Rawls 1967, 59–60.

© The Author(s), 2021. Published by Canadian Journal of Philosophy. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

These moral-aggregation critics object that moral justifications should not be based on comparisons between aggregates of people's claims or well-being.[3] Unfortunately, this objection, which we may call the *Objection from Moral Aggregation*, is rarely put forward in a precise manner. Still, a plausible explication is that the objection rejects justifications that involve moral aggregation in the following sense:[4]

> A justification of a moral evaluation involves *moral aggregation* if and only if the justification is fundamentally based in part on a comparison where at least one of the relata is an aggregate of the claims or well-being of more than one individual.

Rejecting moral aggregation means accepting

> *The Individualist Restriction*: The only comparisons that a justification of a moral evaluation may be fundamentally based on are comparisons where no relatum is an aggregate of the claims or well-being of more than one individual.

The Objection from Moral Aggregation is not that moral evaluations of aggregates of claims are necessarily problematic. What is supposed to be problematic is that comparisons of such aggregates are part of the justifications of moral evaluations. So the evaluation that it's better to save the many than to save the one needn't be problematic. The target of the Objection from Moral Aggregation is the *justification* of this evaluation by the Total Principle or by some other form of moral aggregation.[5] In fact, many moral-aggregation critics believe that there is an adequate justification of its being better to save the many than to save the one.[6] They believe that, while the standard utilitarian justification involves moral aggregation, there is an alternative justification that does not—namely, the Argument for Best Outcomes.[7]

In this paper, I will extend the Argument for Best Outcomes with a further principle to show that *any* utilitarian evaluation can be justified without relying on the Total Principle or any other form of moral aggregation.

---

[3]In taking the problem of moral aggregation to be a problem about justification, I'm following Taurek. He argues that the relative numbers of people involved or any notion of the sum of different people's losses or gains shouldn't be part of the justification of acts and duties (1977, 312), nor a 'ground for a moral obligation' (297–302), nor a 'source or derivation' of duties (310), nor 'something in itself of significance in determining our course of action' (293), nor something 'that has relevance for choice and preference' (2021, 321).

[4]The aim here is to capture the form of moral aggregation which Taurek (1977, 307–10, 313) and Scanlon (1998, 229–30) find problematic in moral justifications. My account is, I think, a better interpretation of what the moral-aggregation critics object to than Hirose's (2015, 24) extensional account. On Hirose's account, no lexical principle for evaluating outcomes would be aggregative. Consider, for example, a variant of utilitarianism that uses the Leximax Equity Criterion (defined later) as a tiebreaker in case two outcomes have the same sum total of well-being. According to this lexical variant of utilitarianism, an outcome X is at least as good as an outcome Y if and only if, and because, either (i) X has a greater sum total of well-being than Y or (ii) the outcomes have equal sum totals of well-being and X would be at least as good as Y according to the Leximax Equity Criterion. This variant seems to involve a form of moral aggregation that's objectionable on the same grounds as the standard utilitarian justification by the Total Principle, but it wouldn't be aggregative on Hirose's account; see Gustafsson 2017, 966–67. On the other hand, Fleurbaey and Tungodden's (2010, 402) Minimal Aggregation condition is satisfied by some plausibly non-aggregative theories such as the *Maximax Equity Criterion*, which says that an outcome X is at least as good as outcome Y if and only if the maximum well-being of any individual is at least as high in X as in Y.

[5]Taurek (2021, 321–22), for example, admits that he has no compelling objection to someone who judges that it's better to save the many than to save the one if this evaluation is not based on (nor unmediated by) the alleged fact that the combined suffering of the many would be greater than the suffering for the one.

[6]Among others, Kamm (1993, 75–98) and Scanlon (1998, 229–41).

[7]See Kamm 1993, 85, where it was called the Aggregation Argument. The new name comes from Kamm 2007, 32. For a structurally similar objection to indifference between saving the one and saving the many (which does not rely on Anonymity), see Kavka 1979, 291–92.

1. The Argument for Best Outcomes

The Argument for Best Outcomes relies on three principles.[8] The first is based on the idea that morality demands impartiality between people, other things being equal (Sen 1974, 391 and Blackorby, Bossert, and Donaldson 2005, 49):

> *Anonymity*: If outcomes X and Y only differ in that the identities of some people who exist in these outcomes have been permuted, then X and Y are equally good.

This principle is sometimes called 'Impartiality'.[9] But the principle requires more than mere impartiality between outcomes that are alike except for a permutation of identities; it requires that the outcomes are equally good. It wouldn't be any less impartial if the outcomes were incomparable in value than if they were equally good. Because, just like equality, incomparability is symmetric. It doesn't favour any one of the relata.

While Anonymity is compelling, it isn't beyond dispute: Anonymity rules out partiality, and partiality is part of common-sense morality (specifically, the idea that you may give extra weight to your own well-being and the well-being of your friends and family).[10] Yet, for the purposes of our current discussion, the key feature of Anonymity is not that it's self-evident or undeniable but that it's free from moral aggregation—that is, Anonymity does not involve any comparisons of aggregates of people's claims or well-being. This feature is still clearer for the following weakened variant, which suffices for the argument:

> *Pairwise Anonymity*: If outcomes X and Y only differ in that the identities of two people who exist in these outcomes have been permuted, then X and Y are equally good.

Consider the following outcomes A and B, which only differ in that the identities of two people (P_1 and P_2) have been permuted (a third person, P_3, is unaffected):

	P_1	P_2	P_3
A	4	0	0
B	0	4	0

Since A and B only differ in that the identities of P_1 and P_2 have been permuted, Pairwise Anonymity entails that A and B are equally good. If two outcomes only differ in that the identities of two people have been permuted, then no further person is affected and any loss for one of the two is perfectly matched by a gain for the other.[11] In a choice between A and B, for instance, any loss for one of P_1

[8] Here, I follow Hirose's (2001, 341) axiomatic presentation of the Argument for Best Outcomes. An advantage of his presentation is that it makes clear that the Argument for Best Outcomes isn't open to Otsuka's (2000, 291–92) objection that the argument implicitly balances aggregates of claims.

[9] See, for instance, Hirose 2001, 341.

[10] A strong argument against partiality is that it leads to outcomes that are worse for all parties in some Prisoner's Dilemma situations; see Parfit 1984, 95–98. For the original Prisoner's Dilemma case, see Tucker 1980, 101.

[11] To see that this needn't be the case with Anonymity, consider the following outcomes (Chapman 2010, 182):

	P_1	P_2	P_3
A'	3	1	2
B'	1	2	3

Outcome B' is just like outcome A' except that people's identities have been permuted. Accordingly, Anonymity entails that A' and B' are equally good. But P_1 loses 2 units of well-being if B' is chosen over A', while no one gains as much. So there's no parity

and P_2 is perfectly matched by a gain for the other. So, by only making one-to-one comparisons between individuals, we have that there is an equivalence of gains and losses between *A* and *B*. Even though this justification balances gains against losses, it only balances the gain for one individual against the loss for another individual. Hence the justification avoids moral aggregation, and it conforms to the Individualist Restriction.[12]

The second principle is based on the idea that if one outcome dominates another outcome in terms of individual well-being, then its better (Broome 1987, 410; 1991, 165):[13]

> *The Strong Principle of Dominance*: If (i) the same people exist in outcomes *X* and *Y*, (ii) each of these people has at least as high well-being in *X* as in *Y*, and (iii) some person has higher well-being in *X* than in *Y*, then *X* is better than *Y*.

Consider the following outcomes *B* and *C*, where everyone is equally well off in *B* as in *C* except P_3 who is better off in *C* than in *B*:

	P_1	P_2	P_3
B	0	4	0
C	0	4	4

By comparing each person's well-being in *B* with their well-being in *C*, we can conclude that each person has at least as high well-being in *C* as in *B* and that P_3 has higher well-being in *C* than in *B*. Based on these intrapersonal comparisons, the Strong Principle of Dominance entails that *C* is better than *B*. This justification does not involve moral aggregation because it doesn't balance claims or well-being between *different* people.

The third principle is the following principle of the logic of value (Arrow 1951, 13; Sen 1970, 2; 2017, 47; and Quinn 1977, 77):

> *Transitivity*: If outcome *X* is at least as good as outcome *Y* and *Y* is at least as good as outcome *Z*, then *X* is at least as good as *Z*.

---

of individual gains and losses between A' and B'. Yet, since more than two people's identities are permuted in the move from A' to B', Pairwise Anonymity does not entail that these outcomes are equally good. To derive that conclusion, we need to apply Pairwise Anonymity twice (for example, permute P_1 and P_2 in A' then permute P_2 and P_3) and then apply Transitivity (defined later).

[12]We may be able to justify Anonymity (and the logically weaker Pairwise Anonymity) without balancing any gains and losses. An alternative justification is based on the claim that personal identities have no moral significance: it's only the list of well-being levels that is of moral concern, not who has which level. On this justification, we don't need to compare any gains or losses to derive that *A* and *B* are equally good. We only need to compare the well-being levels between individuals: P_1, P_2, and P_3 in *A* have the same well-being as P_2, P_1, and P_3 in *B* respectively. A disadvantage of this alternative justification of Anonymity is that it may seem to violate the separateness of persons (see Gauthier 1963, 126–27; Nagel 1970, 138; Rawls 1971, 24; 1999, 27; and Nozick 1974, 32–33). *The Objection from the Separateness of Persons* is, roughly, the objection that losses can only be legitimate if they are compensated whereas a loss for one person cannot be compensated by any gains for other people. Yet it's hard to know what to make of this objection. Many of those who insist on the separateness of persons (for instance, Rawls 1971, 83; 1999, 72 and Nagel 1970, 142; 1978, 22) defend the Difference Principle (see note 17). Yet the Difference Principle also entails (i) Anonymity and (ii) that personal identities do not matter in the sense that it doesn't matter who has which well-being level (see Brink 2020, 386–88). Could the Objection from the Separateness of Persons challenge Pairwise Anonymity? Consider the use of Pairwise Anonymity in, for instance, the justification of A's being equally as good as *B*. If *A* is replaced by *B*, then P_1 suffers an uncompensated loss. But, if *B* is replaced by *A*, then P_2 suffers an equally great uncompensated loss. So, in terms of uncompensated losses, *A* and *B* seem equally bad—and, thus, equally good. Hence the separateness of persons does not challenge Pairwise Anonymity.

[13]The clause that the same people exist in outcomes *X* and *Y* should be read as saying that the set of people who exist in *X* is the same as the set of people who exist in *Y*.

From (i) that *A* and *B* are equally good and (ii) that *C* is better than *B*, it follows by Transitivity that *C* is better than *A*. As long as the first two evaluations—(i) and (ii)—have been justified without moral aggregation, Transitivity provides a justification of *C*'s being better than *A* which does not involve moral aggregation (because, for this justification, Transitivity does not rely on any other comparisons than the first two).

With these principles, we can state the Argument for Best Outcomes. Suppose that getting 4 units of well-being in outcomes *A*, *B*, and *C* corresponds to getting saved and that getting 0 units corresponds to not being saved. In *A*, only P_1 is saved. In *B*, only P_2 is saved. And, in *C*, both P_2 and P_3 are saved but P_1 is not. Hence we have the following outcomes:[14]

	P_1	P_2	P_3
A	4	0	0
B	0	4	0
C	0	4	4

We can then argue as follows:

The Argument for Best Outcomes

(1)	*A* and *B* are equally good.	Pairwise Anonymity
(2)	*C* is better than *B*.	The Strong Principle of Dominance
(3)	*C* is better than *A*.	(1), (2), Transitivity

We have argued, without relying on moral aggregation, that *C* is better than *A*. The difference between *A* and *C* is that, if *A* were chosen over *C*, only one person (P_1) would be saved but, if *C* were chosen, two other people (P_2 and P_3) would be saved. Therefore, we have an argument for its being better that a greater number of people are saved, and this argument does not rely on moral aggregation.[15]

It may be objected that the Argument for Best Outcomes relies on moral aggregation in the move from (1) and (2) to (3). The evaluation in (3) is justified by (1), (2), and Transitivity. So *C*'s being better than *A* is justified in part by *A*'s being equally as good as *B* and in part by *C*'s being better than *B*. But *A*'s being equally as good as *B* is a comparison of the whole of outcome *A* with the whole of outcome *B*. And *C*'s being better than *B* is a comparison of the whole of outcome *C* with the whole of outcome *B*. Each of these compared outcomes includes the well-being of three people. Hence the justification of the evaluation in (3) is based in part on comparisons where at least one of the relata is an aggregate of (among other things) the well-being of more than one individual.

Even so, this does not show that the Argument for Best Outcomes involves moral aggregation, because these comparisons that the justification of (3) is based on—that is, (1) and (2)—can in turn be justified without moral aggregation. So the justification of (3) by (1), (2), and Transitivity is not

[14]Here and in the rest of the paper, we assume that outcomes with all possible distributions of well-being exist. See Broome 1991, 80–81.

[15]While we have applied the Argument for Best Outcomes to a one-versus-two case, the argument also works, changing what needs to be changed, for any *n*-versus-*m* case, where *m* is greater than *n*. Just replace P_1 with the people in the *n*-sized group, replace P_2 with *n* people out of the *m*-sized group, and replace P_3 with the remaining people in the *m*-sized group. The only difference, in case there are two or more people in the *n*-group, is that Pairwise Anonymity is no longer sufficient to justify (1). So, in that case, we need to justify (1) either by Anonymity or by repeated application of Pairwise Anonymity and Transitivity.

fundamentally based on a comparison where at least one of the relata is an aggregate of the well-being of more than one individual.[16]

2. The Extended Argument for Best Outcomes

The Argument for Best Outcomes can support the utilitarian evaluation that saving the greater number is better if the competing claims have equal strength. Still, the three principles that the argument relies on are too weak to allow us to derive all utilitarian evaluations. For instance, these principles are too weak to show that saving the many is better than saving the one if the benefit for the one is greater than the benefit for each of the many. Consider an outcome D where P_2 and P_3 are saved but their well-being is slightly lower than P_1's well-being in outcome A:

	P_1	P_2	P_3
A	4	0	0
D	0	3	3

To see that no valid argument based on just Anonymity, the Strong Principle of Dominance, and Transitivity could show that D is better than A, consider the Leximax Equity Criterion—a variant of the Leximin Equity Criterion which prioritizes the better off rather than the worse off.

The Leximax Equity Criterion evaluates outcomes with the same population as follows: If the best off in a first outcome are better off than the best off in a second outcome, then the first outcome is better than the second outcome. If the best off in the outcomes are equally well-off, remove one of the best off in each outcome and repeat the test until one outcome emerges as better than the other or there is no one left in the outcomes. If there is no one left in the outcomes, then the outcomes are equally good.

The Leximax Equity Criterion satisfies Anonymity, the Strong Principle of Dominance, and Transitivity, but it entails that A is better than D (and thus that D is not better than A), because the best-off person in A is better off than each of the best-off people in D (d'Aspremont and Gevers 1977, 204). Therefore, since utilitarianism entails that D is better than A, there is at least one utilitarian evaluation that cannot be derived with just Anonymity, the Strong Principle of Dominance, and Transitivity.

So, in order to justify the evaluation that D is better than A, we need an additional principle. And, if we want to justify this evaluation without moral aggregation, the additional principle cannot rely on moral aggregation. Even so, there is a principle that fits the bill. Consider

> *Supervenience on Individual Stakes*: If the same people exist in outcomes X, Y, U, and V and, for each person P who exists in these outcomes, P's well-being in X minus P's well-being in Y is equal to P's well-being in U minus P's well-being in V, then X and Y are equally good if and only if U and V are equally good.

This principle says that, if everyone stands to gain or lose the same amount if X were chosen over Y as they would if U were chosen over V, then the evaluation of these pairs should be the same (in the sense that, if the outcomes in one pair are equally good, the outcomes in the other pair should be so

[16]Timmermann (2004, 109n3) objects that, while neither Pairwise Anonymity nor the Strong Principle of Dominance involves moral aggregation, their conjunction does so. Note, however, that the Argument for Best Outcomes does not rely on this conjunction in the justification of any moral evaluation. Claim (1) is justified by Pairwise Anonymity alone, and claim (2) is justified by the Strong Principle of Dominance alone. The conjunction of (1) and (2), which we derive from the conjuncts by propositional logic, is not a further moral evaluation in need of any further moral justification. Hence the conjunction of (1) and (2) need not be justified by the conjunction of Pairwise Anonymity and the Strong Principle of Dominance.

too). Note that the consequent of Supervenience on Individual Stakes is biconditional; it only lets us derive that *X* and *Y* are equally good conditional on that *U* and *V* are equally good (and vice versa). If the evaluation that *U* and *V* are equally good is justified without violating the Individualist Restriction, then Supervenience on Individual Stakes can justify that *X* and *Y* are equally good without violating the Individualist Restriction, because, in addition to the evaluation of *U* and *V*, Supervenience on Individual Stakes only relies on intrapersonal comparisons of gains and losses between pairs of outcomes. Hence, if *U*'s being equally as good as *V* can be justified without moral aggregation, then *X*'s being equally as good as *Y* can be justified by Supervenience on Individual Stakes without relying on moral aggregation.

For an example illustrating the application of Supervenience on Individual Stakes, consider the following pairs of outcomes:

	P_1	P_2	P_3
A	4	0	0
E	2	2	0

	P_1	P_2	P_3
F	2	0	2
G	0	2	2

If outcome *A* were chosen over outcome *E*, then P_1 would be 2 units better off, P_2 would be 2 units worse off, and P_3 would be neither better nor worse off. And, if outcome *F* were chosen over outcome *G*, we get the same result: P_1 would be 2 units better off, P_2 would be 2 units worse off, and P_3 would be neither better nor worse off. Since, in this manner, each individual stands to gain or lose the same amount if *A* were chosen over *E* as they would if *F* were chosen over *G*, Supervenience on Individual Stakes entails that *A* and *E* are equally good if *F* and *G* are equally good. Suppose that the evaluation that *F* and *G* are equally good is justified by Pairwise Anonymity (a justification that doesn't rely on moral aggregation). Then the evaluation that *A* and *E* are equally good can be justified by Supervenience on Individual Stakes without relying on moral aggregation.

The point here is not that Supervenience on Individual Stakes is self evident or uncontroversial. The principle reflects utilitarianism's insensitivity to whether the distribution of well-being is equal, which is controversial from the perspective of some egalitarian theories.[17] While there's no difference with respect to inequality between *F* and *G*, there is more inequality in *A* than in *E*.[18] For the purposes of our discussion, however, the key feature of Supervenience on

[17]To see that Supervenience on Individual Stakes rules out the evaluative version of Rawls's (1971, 83; 1999, 72) Difference Principle, consider the following pairs of outcomes:

	P_1	P_2
H	2	2
I	1	3

	P_1	P_2
J	2	1
K	1	2

The evaluative version of the Difference Principle can be stated as follows:

> *The Evaluative Difference Principle*: Outcome *X* is at least as good as outcome *Y* if and only if the minimum well-being of any person is at least as high in *X* as in *Y*.

According to the Evaluative Difference Principle, we have that outcome *H* is better than outcome *I* and that outcomes *J* and *K* are equally good. But, if *J* and *K* are equally good, we have, from Supervenience on Individual Stakes, that *H* and *I* are equally good.

[18]Moving from *A* to *E* involves a transfer of well-being from a better-off person to a worse-off person (and this transfer does not make the recipient better off than the donor). So, by the Pigou-Dalton principle (Pigou 1912, 24 and Dalton 1920, 351), *E* is more equal than *A*.

Individual Stakes is that it satisfies the Individualist Restriction and hence that it doesn't involve moral aggregation.[19]

We have that each one of Pairwise Anonymity, the Strong Principle of Dominance, Supervenience on Individual Stakes, and Transitivity satisfies the Individualist Restriction. And, with these four principles, we can derive that *D* is better than *A*. Hence we can justify *D*'s being better than *A* without resorting to moral aggregation in any step. To see this, consider once more the following outcomes:

	P_1	P_2	P_3
A	4	0	0
E	2	2	0
F	2	0	2
G	0	2	2
D	0	3	3

We then argue as follows:

The Extended Argument for Best Outcomes

(1) *F* and *G* are equally good.	Pairwise Anonymity
(2) *A* and *E* are equally good.	(1), Supervenience on Individual Stakes
(3) *E* and *G* are equally good.	Pairwise Anonymity
(4) *D* is better than *G*.	The Strong Principle of Dominance
(5) *D* is better than *A*.	(2), (3), (4), Transitivity

Hence we have an argument that it can be better that two people each get a smaller benefit than that one person gets a larger benefit. And, crucially, this argument does not rely on moral aggregation.

3. A justification of utilitarianism without moral aggregation

The Extended Argument for Best Outcomes can be used to defend utilitarianism against the Objection from Moral Aggregation. The argument's four principles jointly entail, as we shall see, the same evaluations as utilitarianism given a fixed population of two or more people. In other words, the four principles of the Extended Argument for Best Outcomes jointly entail a value ranking of any pair of outcomes in which the same (two or more) people exist, and this ranking will

[19]For a further explanation why Supervenience on Individual Stakes doesn't involve moral aggregation, note that Supervenience on Individual Stakes is consistent with (and suggested by) Parfit's (n.d., chap. 6) principle '*Minimax Loss*: The best outcome is the one in which the greatest loser loses least.' We can generalize Parfit's principle as follows (matching the model of the Minimax-Regret Rule in Savage 1951, 59 and Milnor 1954, 50):

> *The Minimax-Loss Principle*: Outcome *X* is at least as good as outcome *Y* if and only if the greatest loss in well-being for any person if *Y* were chosen over *X* is at least as great as the greatest loss in well-being for any person if *X* were chosen over *Y*.

Given the Minimax-Loss Principle, it would be worse if a single person suffers a major loss than if a large number of people each suffers a small loss, other things being equal. This view avoids moral aggregation, yet it entails Supervenience on Individual Stakes. Therefore, Supervenience on Individual Stakes cannot involve moral aggregation.

coincide with the utilitarian value ranking of these outcomes. So, given that there are at least two people, these principles entail a version of utilitarianism which is restricted to evaluations with a fixed population, namely,

> *Fixed-Population Utilitarianism*: If the same people exist in outcomes *X* and *Y*, then *X* is at least as good as *Y* if and only if the sum total of well-being is at least as great in *X* as in *Y*.

Moreover, two of the principles in the Extended Argument for Best Outcomes are stronger than necessary. We can weaken both Transitivity and the Strong Principle of Dominance and still derive Fixed-Population Utilitarianism. Consider the following weakening of Transitivity:[20]

> *Fixed-Population Transitivity*: If (i) the same people exist in outcomes *X*, *Y*, and *Z*, (ii) *X* is at least as good as *Y*, and (iii) *Y* is at least as good as *Z*, then *X* at least as good as *Z*.

And consider the following weakening of the Strong Principle of Dominance:[21]

> *The Weak Principle of Dominance*: If (i) some person exists in outcomes *X* and *Y*, (ii) the same people exist in *X* and *Y*, and (iii) each of these people has higher well-being in *X* than in *Y*, then *X* is better than *Y*.

These weakened principles along with Pairwise Anonymity and Supervenience on Individual Stakes jointly entail the same evaluations as Fixed-Population Utilitarianism for finite populations of at least two people. We can prove the following theorem:[22]

> Given that the total number of people is finite and greater than one, Fixed-Population Utilitarianism is true if and only if the following principles are all true:
>
> - Fixed-Population Transitivity,
> - Pairwise Anonymity,
> - Supervenience on Individual Stakes, and
> - The Weak Principle of Dominance.

From this theorem, we have that each one of Fixed-Population Utilitarianism's evaluations of outcomes with at least two people can be justified by Fixed-Population Transitivity, Pairwise Anonymity, Supervenience on Individual Stakes, and the Weak Principle of Dominance. And,

[20]By weakening Transitivity to fixed-population cases, we avoid some controversial variable-population cases. For example, the mere-addition paradox (see McMahan 1981, 122–23 and Parfit 1982, 158–60) have lured some people, such as Temkin (1987), to reject Transitivity.

[21]You may wonder why clause (i) is needed. Note that without this clause, the Weak Principle of Dominance would be inconsistent with the existence of unpopulated outcomes. Suppose that no people exist in *X* and *Y*. Then clause (ii) holds—see note 13. And, given the convention that universal quantifications over empty domains are vacuously true, clause (iii) holds too (see Gustafsson 2020, 129n40). So we would have that *X* is better than *Y* and that *Y* is better than *X*, which violates the asymmetry of betterness (see Halldén 1957, 25 and Chisholm and Sosa 1966, 247).

[22]See appendix A for proof. For some closely related theorems, see Milnor 1954, 53; d'Aspremont and Gevers 1977, 203; and Blackorby, Bossert, and Donaldson 2002, 569. Note that these earlier theorems, unlike the one presented in this paper, all assume Completeness, which is controversial. (*Completeness* is the principle that outcome *X* is at least as good as outcome *Y* or *Y* is at least as good as *X*. See Chang 1997 for an overview of the chief worries about Completeness.) Hence the new theorem has an advantage over these earlier theorems. But, for the main argument in this paper, this difference between these theorems doesn't matter much, because Completeness doesn't involve moral aggregation. Another difference is that my proof relies on Pairwise Anonymity rather than Anonymity. While this difference is mathematically trivial, it helps my argument that utilitarianism doesn't rely on moral aggregation, since—as we saw in note 11—it is more obvious that Pairwise Anonymity avoids moral aggregation than that Anonymity does so. Moreover, Pairwise Anonymity has the same advantage over Denicolò's (1999, 276–77) strengthened variant of Anonymity that allows him to drop Transitivity in his characterization of utilitarianism.

since none of these principles involves moral aggregation, this justification of Fixed-Population Utilitarianism does not violate the Individualist Restriction.[23] So utilitarianism can sidestep the Objection from Moral Aggregation.[24]

To derive the same evaluations as utilitarianism for fixed populations with fewer than two people, we also need the following principle of the logic of value (Arrow 1951, 14; Chisholm and Sosa 1966, 248; and Sen 1970, 2; 2017, 47):[25]

Reflexivity: Outcome X is at least as good as X.

Reflexivity does not involve moral aggregation. It just compares an outcome with itself. So there are no relevant claims of any individual. We can prove the following corollary:[26]

Given that the total number of people is finite, Fixed-Population Utilitarianism is true if and only if the following principles are all true:

- Fixed-Population Transitivity,
- Pairwise Anonymity,
- Reflexivity,
- Supervenience on Individual Stakes, and
- The Weak Principle of Dominance.

But, since we only need Reflexivity to evaluate outcomes with fewer than two people, this corollary won't matter for our discussion of moral aggregation. Moral aggregation requires at least two people.

It may be objected that, if we were to justify utilitarian evaluations with these non-aggregative principles, we would still end up with extensionally the same evaluations as if we evaluated outcomes on the basis of their sum total of well-being. So we would still evaluate *as if* we evaluated on the basis of moral aggregation. But, first, note that we would also evaluate as if we didn't evaluate

[23]If we replace *outcome* with *prospect* and *well-being* with *expected well-being* in these principles, we can justify the following subjective version of utilitarianism in the same manner without moral aggregation:

Subjective Fixed-Population Utilitarianism: If the same people exist in prospects X and Y, then X is at least as good as Y if and only if the sum total of expected well-being is at least as great in X as in Y.

[24]Note that not just any characterization of utilitarianism will do for this purpose, because many such characterizations include a principle that seems to involve some form of moral aggregation. Still, the proofs in Maskin 1978, 94 and Blackorby, Bossert, and Donaldson 2005, 118 should also work (but with the slight drawback of some more complicated conditions). Harsanyi's (1955, 313–14) social-aggregation theorem could also work, but it requires that we assume the axioms of Expected Utility Theory for the (personal and impersonal) value orderings of risky prospects. These axioms don't allow for objective versions of utilitarianism (see Gustafsson 2019, 194–195 for sources) which take the value of a prospect to be equal to the value of the outcome that would be the final outcome of the prospect. This violates two axioms of Expected Utility Theory, namely, Independence (Jensen 1967, 173) and Continuity (von Neumann and Morgenstern 1944, 26–27 and Blackwell and Girshick 1954, 106). Moreover, Harsanyi's proof requires that we assume Completeness (see note 22) for both personal and impersonal value orderings. This, as Broome (1987, 418; 2015, 258–59) points out, is a significant drawback. Notably, the characterization in this paper does not require Completeness.

[25]If we allow outcomes without people, then Reflexivity conflicts with *Average Utilitarianism*, the view that an outcome X is at least as good as an outcome Y if and only if the average well-being is at least as high in X as in Y. (See, for example, Harsanyi 1955, 316.) Since there is no well-defined average of well-being for outcomes without people, Average Utilitarianism entails that an outcome without people is not at least as good as itself. (The Evaluative Difference Principle—see note 17—and the Minimax-Loss Principle—see note 19—yield much the same problem.) To get around this problem, we could replace Reflexivity with *Populated Reflexivity*, the principle that, if some person exists in outcome X, then X is at least as good as X. Given that we replace Reflexivity with Populated Reflexivity and assume that the total number of people is not only finite but also greater than zero, the corollary will still hold.

[26]See appendix B for proof.

on the basis of moral aggregation, since we would also evaluate as if we merely applied the above principles. And, second, note that, however we evaluate outcomes, there will always be a way of justifying an extensionally equivalent evaluation of outcomes on the basis of some (perhaps convoluted) form of moral aggregation. Hence, on the one hand, if the Objection from Moral Aggregation is that we shouldn't evaluate *as if* we evaluated on the basis of moral aggregation, it seems to prove too much, since it would rule out any way of evaluating outcomes. On the other hand, if the objection is merely that moral aggregation shouldn't figure in the justification of evaluations, then it shouldn't cause concern about utilitarianism, since, by way of the above principles, the utilitarian evaluations can be justified without moral aggregation.

Acknowledgments. I wish to thank Gustaf Arrhenius, Krister Bykvist, Bruce Chapman, Richard Yetter Chappell, Tomi Francis, Bernward Gesang, Christopher Jay, Kacper Kowalczyk, Mary Leng, Alastair Norcross, Martin Peterson, Christian Piller, Mozaffar Qizilbash, Wlodek Rabinowicz, Daniel Ramöller, Korbinian Rueger, Dean Spears, Tom Stoneham, Frans Svensson, Elliott Thornley, John A. Weymark, the audience at ISUS 2018, Karlsruhe Institute of Technology (26 July 2018), and two anonymous referees for valuable comments. Financial support from the Swedish Foundation for Humanities and Social Sciences is gratefully acknowledged.

Johan E. Gustafsson is an associate professor at University of Gothenburg and Institute for Futures Studies and a senior researcher at University of York. He is currently at work on the book *Money-Pump Arguments*, which is forthcoming from Cambridge University Press.

References

Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. New York: Wiley.

Blackorby, Charles, Walter Bossert, and David Donaldson. 2002. "Utilitarianism and the Theory of Justice." In *Handbook of Social Choice and Welfare*, vol. 1, edited by Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzamura, 543–96. Amsterdam: Elsevier.

Blackorby, Charles, Walter Bossert, and David Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*, Cambridge: Cambridge University Press.

Blackwell, David, and M. A. Girshick. 1954. *Theory of Games and Statistical Decisions*. New York: Wiley.

Brink, David O. 2020. "Consequentialism, the Separateness of Persons, and Aggregation." In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore, 378–400. New York: Oxford University Press.

Broome, John. 1987. "Utilitarianism and Expected Utility." *The Journal of Philosophy* 84 (8): 405–22.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Broome, John. 2015. "General and Personal Good: Harsanyi's Contribution to the Theory of Value." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 249–66. New York: Oxford University Press.

Chang, Ruth. 1997. "Introduction." In *Incommensurability, Incomparability, and Practical Reason*, edited by Ruth Chang, 1–34, Cambridge, MA: Harvard University Press.

Chapman, Bruce. 2010. "Preference, Pluralism, and Proportionality." *University of Toronto Law Journal* 60 (2): 177–96.

Chisholm, Roderick M., and Ernest Sosa, 1966. "On the Logic of Intrinsically Better." *American Philosophical Quarterly* 3 (3): 244–49.

Dalton, Hugh. 1920. "The Measurement of the Inequality of Incomes." *The Economic Journal* 30 (119): 348–61.

d'Aspremont, Claude and Louis Gevers. 1977. "Equity and the Informational Basis of Collective Choice." *The Review of Economic Studies* 44 (2): 199–209.

Denicolò, Vincenzo. 1999. "A Characterization of Utilitarianism without the Transitivity Axiom." *Social Choice and Welfare* 16 (2): 273–78.

Fleurbaey, Marc, and Bertil Tungodden. 2010. "The Tyranny of Non-Aggregation versus the Tyranny of Aggregation in Social Choices: A Real Dilemma." *Economic Theory* 44 (3): 399–414.

Gauthier, David P. 1963. *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and their Exemplification in Discourse*. Oxford: Clarendon Press.

Gustafsson, Johan E. 2017. "Review of Iwao Hirose, *Moral Aggregation*." *Mind* 126 (503): 964–67.

Gustafsson, Johan E. 2019. "Is Objective Act Consequentialism Satisfiable?" *Analysis* 79 (2): 193–202.

Gustafsson, Johan E. 2020. "Permissibility Is the Only Feasible Deontic Primitive." *Philosophical Perspectives* 34 (1): 117–33.

Halldén, Sören. 1957. *On the Logic of Better*. Lund: C. W. K. Gleerup.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *The Journal of Political Economy* 63 (4): 309–21.

Hirose, Iwao. 2001. "Saving the Greater Number without Combining Claims." *Analysis* 61 (4): 341–42.

Hirose, Iwao. 2015. *Moral Aggregation*. New York: Oxford University Press.
Jensen, Niels Erik. 1967. "An Introduction to Bernoullian Utility Theory: I. Utility Functions." *Swedish Journal of Economics* 69 (3): 163–83.
Kamm, F. M. 1993. *Morality, Mortality Volume I: Death and Whom to Save from It*. New York: Oxford University Press.
Kamm, F. M. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.
Kavka, Gregory S. 1979. "The Numbers Should Count." *Philosophical Studies* 36 (3): 285–94.
Maskin, Eric. 1978. "A Theorem on Utilitarianism." *The Review on Economic Studies* 45 (1): 93–96.
McMahan, Jeff. 1981. "Problems of Population Theory." *Ethics* 92 (1): 96–127.
Milnor, John. 1954. "Games against Nature." In *Decision Processes*, edited by Robert M. Thrall, Clyde H. Coombs, and Robert L. Davis, 49–59. New York: Wiley.
Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.
Nagel, Thomas. 1978. "The Justification of Equality." *Crítica: Revista Hispanoamericana de Filosofía* 10 (28): 3–31.
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
Otsuka, Michael. 2000. "Scanlon and the Claims of the Many versus the One." *Analysis* 60 (3): 288–93.
Parfit, Derek. 1982. "Future Generations: Further Problems." *Philosophy & Public Affairs* 11 (2): 113–72.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
Parfit, Derek. n.d. *On Giving Priority to the Worse Off*. Unpublished manuscript.
Pigou, A. C. 1912. *Wealth and Welfare*. London: Macmillan.
Portmore, Douglas W. 2020. "Introduction." In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore, 1–21. New York: Oxford University Press.
Quinn, Philip L. 1977. "Improved Foundations for a Logic of Intrinsic Value." *Philosophical Studies* 32 (1): 73–81.
Rawls, John. 1967. "Distributive Justice." In *Philosophy, Politics, and Society*, Third Series, edited by Peter Laslett and W. G. Runciman, 58–82. Oxford: Blackwell.
Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rawls, John. 1999. *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press.
Savage, Leonard J. 1951. "The Theory of Statistical Decision." *Journal of the American Statistical Association* 46 (253): 55–67.
Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
Sen, Amartya. 1974. "Informational Bases of Alternative Welfare Approaches: Aggregation and Income Distribution." *Journal of Public Economics* 3 (4): 387–403.
Sen, Amartya. 2017. *Collective Choice and Social Welfare: An Expanded Edition*. Cambridge, MA: Harvard University Press.
Taurek, John M. 1977. "Should the Numbers Count?" *Philosophy & Public Affairs* 6 (4): 293–316.
Taurek, John M. 2021. "Reply to Parfit's 'Innumerate Ethics'." In *Principles and Persons: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 311–22. Oxford: Oxford University Press.
Temkin, Larry S. 1987. "Intransitivity and the Mere Addition Paradox." *Philosophy & Public Affairs* 16 (2): 138–87.
Timmermann, Jens. 2004. "The Individualist Lottery: How People Count, but Not Their Numbers." *Analysis* 64 (2): 106–12.
Timmons, Mark. 2013. *Moral Theory: An Introduction*. 2nd ed. Lanham, MD: Rowman & Littlefield.
Tucker, A. W. 1980. "A Two-Person Dilemma." *The UMAP Journal* 1 (1): 101.
von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Appendices

A. Proof of the theorem

We shall prove the theorem that, given that the total number of people is finite and greater than one, Fixed-Population Utilitarianism is true if and only if Fixed-Population Transitivity, Pairwise Anonymity, Supervenience on Individual Stakes, and the Weak Principle of Dominance are all true.[27]

We begin with the right-to-left direction of the biconditional. Suppose that X and Y are outcomes with the same people $P_1, \ldots, P_n$. Consider, first, the case where X and Y have the same sum total of well-being. Starting with this pair of outcomes, we will consider a sequence of pairs of outcomes where, in each pair, the outcomes are equally good if the outcomes in the next pair are equally as

[27]The proof technique is essentially the same as in Milnor's (1954, 53) characterization of the Laplace criterion.

good as each other, until we get to a pair of outcomes that we can show are equally as good as each other.

(SORT): Perform the following sorting procedure on each outcome in the pair: as long as it is not the case, for each $i = 1, \ldots, n - 1$, that P_i has at least as high well-being as P_{i+1} in the outcome, find the smallest integer j such that P_{j+1} has higher well-being than P_j in the outcome and replace the outcome with an outcome that only differs in that the identities of P_j and P_{j+1} have been permuted. It follows, from Pairwise Anonymity, that each new outcome is equally as good as the outcome it replaces. And we have, from Fixed-Population Transitivity, that the resulting sorted outcome is equally as good as the one we started with. Since there are only a finite number of people in the outcomes, this procedure will provide, in a finite number of iterations, a new pair of outcomes with people ordered $(P_1, \ldots, P_n)$ by decreasing well-being. And the outcomes in this new pair are equally good if and only if the outcomes in the previous pair are equally good.

(DECREASE): Then, with the resulting pair of outcomes with people ordered by decreasing well-being, replace those outcomes by two new outcomes that only differ from the old two respectively in that each person's well-being is decreased by whichever is lower of that person's well-being levels in the old pair of outcomes. We have, by Supervenience on Individual Stakes, that the outcomes in the new pair are equally good if and only if the outcomes in the old pair are equally good.

Repeat step SORT followed by step DECREASE until, after a finite number of iterations of these steps, we have a pair of outcomes in which everyone has zero well-being. To see that this is what we'll end up with, note that we started with two outcomes with an equal sum total of well-being and, after SORT or DECREASE, we still have two outcomes with an equal sum total of well-being since SORT leaves the sum totals of well-being unchanged and DECREASE subtracts the same amount of well-being from both outcomes. After the first iteration of DECREASE, any negative well-being has been cancelled out. From then on, since the outcomes have the same sum total of well-being, there are people with positive well-being in one of the outcomes if and only if there are people with positive well-being in the other outcome. Hence, after each further iteration of SORT, there must be at least one person (specifically, P_1) that has positive well-being in both outcomes. In the next iteration of DECREASE, this person will then get their well-being decreased by the lowest of their well-being levels in the two outcomes and thereby end up with zero well-being in at least one of the outcomes. So, with each iteration of DECREASE after the first one, we have that one of the outcomes will have at least one further person with zero well-being. Moreover, since all negative well-being has been cancelled out, DECREASE leaves all people with zero well-being as they are. And SORT leaves the number of people with zero well-being unchanged. Hence, with each further iteration of DECREASE, we'll get more and more people with zero well-being in the outcomes. So, after a finite number of iterations of SORT and DECREASE, we end up with a pair of outcomes X' and Y' where everyone has zero well-being.

Then, let X'' be an outcome that is just like X' except that the identities of P_1 and P_2 have been permuted. By Pairwise Anonymity, we have that X' and X'' are equally good. For each person in these outcomes, the difference in their well-being between X' and Y' is the same as the difference in their well-being between X' and X''—namely, no difference at all. So, by Supervenience on Individual Stakes, we have that X' and Y' are equally good, since X' and X'' are equally good.

Since the outcomes in the final pair are equally good (that is, X' and Y' are equally good), we have that, in each pair in the sequence of pairs we have considered, the outcomes are equally good. Thus we can conclude that the outcomes in the pair we started with are equally good—that is, X and Y are equally good. So we have that, if the sum total of well-being is the same in X and Y, then X and Y are equally good.

We now turn to the case where the sum total of well-being is greater in one of the outcomes. So suppose now that the sum total of well-being is greater in X than in Y. And, as before, suppose that the same people exist in X and Y. Let X' and Y' be two outcomes such that (i) the same people exist in X, Y, X', and Y', (ii) X' has the same sum total of well-being as X, (iii) Y' has the same sum total of well-being as Y, and (iv) each of X' and Y' is perfectly equal—that is, in each of these outcomes, everyone has the same level of well-being. Hence we have that the same people exist in X' and Y' and

that each of these people has higher well-being in X' than in Y'. Then, from the Weak Principle of Dominance, we have that X' is better than Y'. Since X and X' have the same sum total of well-being, we have, by our previous result, that X and X' are equally good. And, since Y and Y' have the same sum total of well-being, we similarly have that Y and Y' are equally good. Then, from Fixed-Population Transitivity, we have that X is better than Y.

So, combining these results, we have that X is at least as good as Y if and only if the sum total of well-being is at least as great in X as in Y. Hence, if Fixed-Population Transitivity, Pairwise Anonymity, Supervenience on Individual Stakes, and the Weak Principle of Dominance are all true, then Fixed-Population Utilitarianism is true given a finite population of at least two people. The second part of the proof—the proof of the biconditional's left-to-right direction—is trivial.

B. Proof of the corollary

We shall prove the corollary that, given that the total number of people is finite, Fixed-Population Utilitarianism is true if and only if Fixed-Population Transitivity, Pairwise Anonymity, Reflexivity, Supervenience on Individual Stakes, and the Weak Principle of Dominance are all true.

We begin with the right-to-left direction of the biconditional. Given the theorem we proved in appendix A, we only need to cover outcomes with fewer than two people. First, we will consider the case where X and Y have the same sum total of well-being.

Suppose that only one person exists in X and Y. We have, by Reflexivity, that X is equally as good as X. And, for the person in X and Y, the difference in their well-being between X and Y is the same as the difference in their well-being between X and X—namely, no difference at all. So, by Supervenience on Individual Stakes, we have that X is equally as good as Y since X is equally as good as X.

Next, suppose that no people exist in X and Y. We have, by Reflexivity, that X is equally as good as X. Since no people exist in X and Y, we (trivially) have that the same people exist in these outcomes.[28] Then, by Supervenience on Individual Stakes, we have that X and Y are equally good.

Finally, we turn to the case where the sum total of well-being is greater in X than in Y. In this case, there has to be one person who exists in X and Y and who has higher well-being in X than in Y. Therefore, since there is only one person in X and Y and this person has higher well-being in X than in Y, we have, from the Weak Principle of Dominance, that X is better than Y.

Hence, if Fixed-Population Transitivity, Pairwise Anonymity, Reflexivity, Supervenience on Individual Stakes, and the Weak Principle of Dominance are all true, then Fixed-Population Utilitarianism is true given a finite population. As before, the proof of the biconditional's left-to-right direction is trivial.

[28]See note 13.

Cite this article: Gustafsson, J. E. 2021. Utilitarianism without Moral Aggregation. *Canadian Journal of Philosophy* 51: 256–269, doi:10.1017/can.2021.20